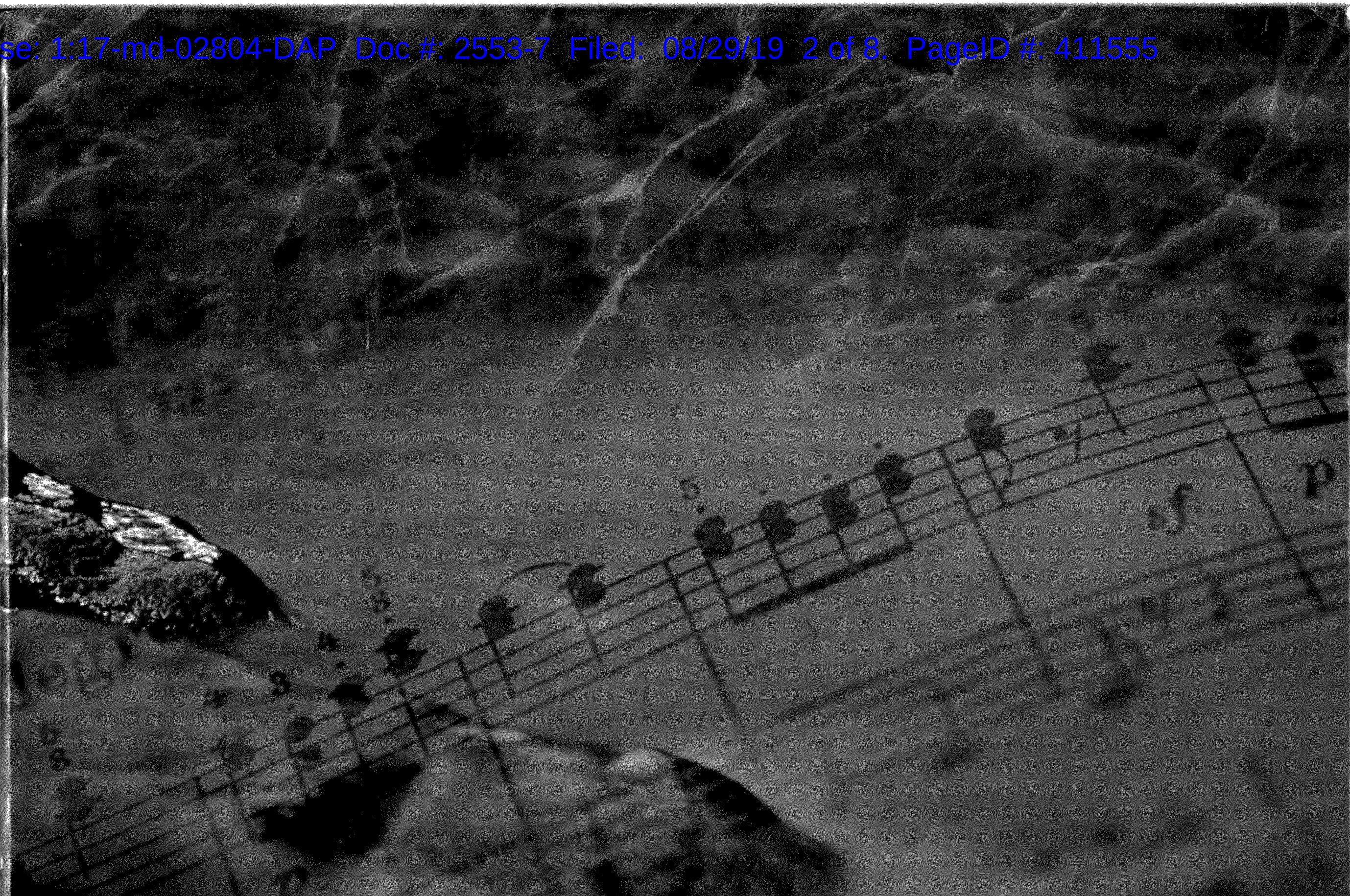
Exhibit 7



ECONOMETRICS

A MODERN INTRODUCTION

MICHAEL P. MURRAY

For Rosanne, with all my love.

Publisher: Daryl Fox
Editor-in-Chief: Denise Clinton
Acquisitions Editor: Adrienne D'Ambrosio
Assistant Editor: Amy Fleischer
Editorial Assistant: Ashley Booth
Managing Editor: Nancy Fenton
Senior Production Supervisor: Katherine Watson
Designer: Leslie Haimes, Nesbitt Graphics, Inc.
Cover Designer: Leslie Haimes
Supplements Supervisor: Kirsten Dickerson
Executive Media Producer: Michelle Neil
Media Producer: Bridget Page
Executive Marketing Manager: Stephen Frail
Marketing Coordinator: Kate MacLean
Rights and Permissions Advisor: Shannon Barbe
Manufacturing Buyer: Carol Melville
Production Coordination: Marilyn Dwyer, Nesbitt Graphics, Inc.
Project Manager: Bonnie Boehme, Nesbitt Graphics, Inc.
Composition: Nesbitt Graphics, Inc.
Illustrations: Nesbitt Graphics, Inc.
Cover photo: © Photonica

Library of Congress Cataloging-in-Publication Data
Murray, Michael P., 1946–
Econometrics : a modern introduction / Michael Murray.
p. cm.
Includes bibliographical references and index.
ISBN 0-321-11361-6 (alk. paper)
1. Econometrics. I. Title.
HB139.M877 2006
330'.01'5195--dc22 2005018586

2 3 4 5 6 7 8 9 10-CRW-09 08 07 06

some variable. When we have just enough instruments for consistent estimation, we say that the regression equation is **exactly identified**. When we have more than enough instruments, we say that the equation is **overidentified**. If the model cannot be consistently estimated because we have too few instruments, we say that the equation is **underidentified**. Because the instruments for troublesome variables cannot themselves be explanators in the model, identification requires that our model must exclude from the regression equation at least as many potential instrumental variables as there are troublesome explanators. We call this requirement the **order condition for identification**.

The order condition is a necessary condition for the identification of an equation.

13.3 Sources of Contemporaneous Correlation

WHAT IS THE DGP?

We have just seen that mismeasured explanators make OLS biased and inconsistent and that IV estimation consistently estimates the parameters of the model in this case. This section explains how IV estimation can provide consistent estimates in three other often encountered settings in which contemporaneously correlated explanators and disturbances make OLS biased and inconsistent. The three settings are (i) omitted relevant explanators, (ii) lagged dependent variables coupled with serially correlated disturbances, and (iii) jointly determined dependent and independent variables.

Instrumental Variables and Omitted Variables Bias

When relevant explanatory variables are excluded from the model, they become a part of the disturbance term. If

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i,$$

but we mistakenly specify

$$Y_i = \beta_0 + \beta_1 X_{1i} + \eta_i,$$

then $\eta_i = \beta_2 X_{2i} + \varepsilon_i$. We learned in Chapter 8 that omitted variables bias arises when the included variables are correlated with the omitted relevant variables, that is, when $\mathrm{E}(x_{1i}x_{2i}) \neq 0$. In such cases, the included explanators are not contemporaneously uncorrelated with the disturbances, $\mathrm{E}(x_{1i}\eta_i) = \mathrm{E}(x_{1i}(\beta_2 X_{2i} + \varepsilon_i)) \neq 0$, so OLS suffers from the bias and inconsistency that we call omitted variables bias.

Notice that in the omitted variable case, our goal is to estimate the coefficient on the regressor included in our model, X_{1i} itself. This is in contrast to the

mismeasured explanator case, when our interest is in the effect of the actual variable, not the effect of the regressor, M_i. In the present case, then, the IV estimator is

$$\beta_{IV} = \frac{(\sum Z_i Y_i)}{(\sum Z_i X_{1i})},$$

because X_1 is itself the troublesome variable.

IV estimation can overcome omitted variables bias if the instruments are valid ones, that is, if they satisfy our two conditions: (i) they are correlated with the troublesome variable; and (ii) they are uncorrelated with the disturbance. For example, Josh Angrist in his study of the effect of military service on earnings, and Mark McClellan and his coauthors in their study of the efficacy of heart disease treatments, use IV estimation to overcome omitted variables bias.

Satisfying requirement (ii) is often particularly difficult when seeking instruments to overcome omitted variables bias. Omitted variable bias arises because the included X's are correlated with the omitted variables. Valid instruments in this case must be variables that are correlated with the included X's and uncorrelated with the omitted X's. A difficult problem in choosing instruments to overcome omitted variable bias is establishing that the instruments are uncorrelated with the omitted variables. McClellan and his coauthors confronted this issue directly. In their heart treatment paper, they conduct formal tests of their success in choosing a suitable instrumental variable. In contrast, Angrist offered only intuitive support for his choice of a lottery "winning" birthday dummy variable as an instrument. How, he asked, could the randomized lottery outcome be correlated with traits of the individuals? Angrist's persuasive rationale is the rhetorical and substantive foundation of the randomized experiments that are often conducted in both the natural and the social sciences.

Instrumental Variables Estimation and Lagged Dependent Variables

Lagged dependent variables are another potential destroyer of the consistency of OLS. Robert Barro's classic money growth equations, discussed in Chapter 6, contained money growth lagged once and lagged twice. We learn here that if the disturbances are serially correlated, OLS may be inconsistent in models with lagged dependent variables.

To see that when lagged dependent variables are coupled with serially correlated disturbances, OLS may become inconsistent, suppose the dependent variable lagged once appears as an explanator in a DGP with first-order autoregressive disturbances:

$$Y_t = \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t$$
$$\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t. \qquad 13.11$$

We can rewrite Y_t, substituting for Y_{t-1}. This exposes the appearance of ε_{t-1} within the explanators:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 Y_{t-1} + \varepsilon_t \\ &= \beta_0 + \beta_1(\beta_0 + \beta_1 Y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t. \end{aligned} \qquad \textbf{13.12}$$

The appearance of ε_{t-1} in Equation 13.12 makes it clear that OLS is inconsistent in this DGP because Equation 13.11 ensures that $E(\varepsilon_t\varepsilon_{t-1})$ is not zero; the explanator Y_{t-1} is contemporaneously correlated with the disturbance, ε_t. *A lagged dependent variable among the explanators and serially correlated disturbances often combine together to make OLS inconsistent.* Serial correlation and lagged dependent variables *may* not make OLS inconsistent, however. For example, suppose that instead of first-order serial correlation, the DGP has a serial correlation in which $E(\varepsilon_t\varepsilon_{t-1}) = 0$, but $E(\varepsilon_t\varepsilon_{t-2}) \neq 0$. In such a case, OLS would be consistent. (Barro found no serial correlation in his disturbances, so he was spared these concerns about the consistency of OLS.)

Notice that in the present case, our goal is to estimate the coefficient on the regressor in our model, Y_{t-1}. This differs from the mismeasured explanator case, in which our interest is in the effect of the true independent variable, not the effect of the regressor, M_i. In the present case, then, the IV estimator is

$$\beta_{IV} = \frac{(\Sigma Z_t Y_t)}{(\Sigma Z_t Y_{t-1})},$$

because Y_{t-1} is the troublesome variable.

IV estimation can consistently estimate the parameters of a model with lagged dependent variables when the disturbances are serially correlated. Finding suitable instruments for the lagged dependent variables is frequently difficult, however. If an explanator besides the lagged dependent variable is present in the model, econometricians sometimes use lagged values of that explanator as an instrument for the lagged dependent variable. Unfortunately, lagged explanators are frequently highly collinear with their current values. In such cases, IV estimators have a high variance for the same reason that OLS estimators have a high variance when explanators are highly collinear.

Instrumental Variables and Jointly Determined Variables

Many economic variables are interdependent. For example, the price of an audio CD influences the quantity of CDs bought and sold, but the quantity produced for sale also influences the CD's price. Similarly, the number of people in jail influences crime rates, but crime rates also influence the number of people in jail. OLS is often biased when used to estimate regression models involving such interdependent variables.

To understand why OLS is biased in such regressions, return to the demand curve for a staple agricultural commodity, wheat, that we examined in Chapter 12. Rather than making the expected quantity demanded strictly proportional to the price of wheat, as we did in Chapter 12, consider a richer specification of the demand for wheat,

$$Q_i = \beta_0 + \beta_1 P_i + \beta_2 I_i + \varepsilon_i^d,$$

in which Q is the quantity of wheat demanded, P is the price of wheat, and I is income, and consider the supply curve for wheat:

$$P_i = \alpha_0 + \alpha_1 Q_i + \alpha_2 W_i + \varepsilon_i^s,$$

where W is an indicator of weather conditions. When there is an above-average quantity of wheat demanded because the demand disturbance, ε_i^d, is greater than zero, price will tend to be higher than if the shock to demand were zero—the higher quantity tends to move farmers up their supply curves. As a consequence, price is contemporaneously correlated with the disturbances in the demand curve, and OLS is biased if it is used to estimate the demand curve. This bias is called **simultaneity bias** because it stems from price and quantity being simultaneously determined.

Fortunately, there is a natural instrumental variable for the troublesome price variable in the demand equation. Weather meets the two requirements for a valid instrument. It is correlated with price because it appears in the supply curve and is also unlikely to be correlated with the shocks in the demand for wheat. IV estimation can sometimes consistently estimate the parameters of a model that contains explanators that are jointly determined with the dependent variables, even though OLS cannot. We call explanators that are jointly determined with the dependent variables **endogenous variables** because they are determined within the system of equations. In this example, price is an endogenous variable that appears as an explanator in the demand equation. Weather, on the other hand, is an example of what we call **exogenous variables** because it is determined outside this system of equations. Exogenous variables are uncorrelated with the disturbances of the system, so they are good candidates for use as instruments. Weather is a good instrument for the demand equation because it does not appear as an explanator in that equation.

OLS applied to the supply equation also suffers from simultaneity bias. In the supply equation, it is quantity that is the troublesome endogenous explanator. Weather is not a useful instrument for quantity in the supply equation because weather already appears in that equation as an explanator. Fortunately, there is a natural instrument for quantity in the supply equation—the income variable that appears in the demand equation, but not in the supply equation. Income is plausi-

bly determined outside the supply-and-demand system for wheat and is therefore plausibly an exogenous variable.

Suitable instruments are not always available when variables are jointly determined. When fewer exogenous variables are excluded from an equation than there are endogenous explanators in the equation, the order condition for identification fails and the equation's coefficients cannot be consistently estimated. Even when enough exogenous variables are excluded from an equation to satisfy the order condition, sometimes no consistent estimator of an equation is possible. Each equation in a system of equations needs its own independent exclusion restrictions. When an equation satisfies the order condition, and the equations' exclusion restrictions are not redundant with the restrictions of other equations in the model, we say that the equation satisfies the **rank condition for identification.** *The rank condition is a sufficient condition for the identification of an equation.*

Lagged Dependent Variables and Identification

We have seen that explanators that are lagged dependent variables make OLS biased and inconsistent if the disturbances are serially correlated. When the disturbances are not serially correlated, however, lagged dependent variables can facilitate consistent estimation of an equation. When there is no serial correlation in the disturbances, lagged dependent variables are uncorrelated with the disturbances. Consequently, if a lagged dependent variable is correlated with a troublesome variable, it can serve as an instrumental variable. Econometricians call a model's exogenous and lagged dependent variables the model's **predetermined variables.**

13.4 An Application: Wage Equations and IV Estimation

We have looked at wage equations several times. One striking feature of wage equations is the many variables they do *not* contain. We explain wages with years of education, age and age squared (or better, experience), sex, race, and occasionally a measure of intelligence. Many other traits do not get measured and therefore do not get included. Did a person's parents read to them, restrict television viewing, send them to good schools, or make them do their homework? All these variables might influence one's future wages, but we do not usually include them in wage equations, because we usually don't measure them. Do these omitted traits bias the OLS estimates we obtain from our wage equations?

Instrumental variables is one strategy for overcoming the omitted variables problem in wage equations. Another strategy is to look at the differences in earnings